

Adaptive quantum state tomography improves accuracy quadratically

D.H. Mahler,¹ Lee A. Rozema,¹ Ardavan Darabi,¹ Chris Ferrie,² Robin Blume-Kohout,³ and A.M. Steinberg¹

¹Centre for Quantum Information & Quantum Control and Institute for Optical Sciences,
Dept. of Physics, 60 St. George St., University of Toronto, Toronto, Ontario, Canada M5S 1A7*
²Institute for Quantum Computing and Department of Applied Mathematics, University of Waterloo
³Sandia National Laboratories

(Dated: March 5, 2013)

We introduce a simple protocol for adaptive quantum state tomography, which reduces the worst-case infidelity ($1 - F(\rho, \hat{\rho})$) between the estimate and the true state from $O(1/\sqrt{N})$ to $O(1/N)$. It uses a single adaptation step and just one extra measurement setting. In a linear optical qubit experiment, we demonstrate a full order of magnitude reduction in infidelity (from 0.1% to 0.01%) for a modest number of samples ($N \approx 3 \times 10^4$).

PACS numbers: 42.50.Dv, 42.50.Xa

Quantum information processing requires reliable, repeatable preparation and transformation of quantum states. *Quantum state tomography* is used to identify the density matrix ρ that was prepared by such a process. No finite ensemble of N samples is sufficient to *uniquely* identify ρ , so we *estimate* it, reporting either a single state $\hat{\rho}$ that is “close” to ρ with high probability [1–5], or a confidence region of nonzero radius that contains ρ with high probability [6, 7]. Both approaches must accept some inaccuracy (the discrepancy between $\hat{\rho}$ and ρ) or imprecision (the diameter of the confidence region). This can be quantified in many ways (e.g., trace norm, fidelity, relative entropy, etc.), but the universal goal of state tomography is to minimize it. In this paper, we consider a well-motivated (see concluding discussion) and popular measure of a point estimator’s inaccuracy, *quantum infidelity*,

$$1 - F(\hat{\rho}, \rho) = 1 - \text{Tr}(\sqrt{\sqrt{\rho}\hat{\rho}\sqrt{\rho}})^2, \quad (1)$$

and its scaling as $N \rightarrow \infty$.

First, we show that the infidelity of standard tomography with *static* measurements can’t beat $1 - F = O(1/\sqrt{N})$ as $N \rightarrow \infty$ for a large and important class of states. Second, we introduce a simple *adaptive* protocol that achieves $1 - F = O(1/N)$ for every state, and explain why it works. Finally, we demonstrate this effect in a linear optical experiment, and achieve a 10-fold improvement in infidelity (from 0.1% to 0.01% with $N = 3 \times 10^4$ measurements) over standard tomography.

Adaptivity has been proposed in various contexts (e.g., Ref. [8] recently treated state estimation as parameter estimation, obtaining a result complementary, but largely orthogonal, to those reported here). Single-step adaptive tomography was first analyzed by [9], and later refined in [10–12]. Here, we present a simple, self-contained derivation of: (1) why quantum fidelity is significant; (2) why adaptive tomography achieves far better infidelity; and (3) how the adaptation should be done. In contrast to prior work, we optimize *worst-case* infidelity, rather

than *average* infidelity. This allows us to achieve high accuracy for *all* states, whereas previous approaches (e.g. [11]) yielded low accuracy on a small but important set of states (see concluding discussion).

ADAPTIVE TOMOGRAPHY

Static tomography uses data from a fixed set of measurements. Different measurements yield subtly different tomographic accuracy [13], but to leading order, “good” protocols for single-qubit tomography provide equal information [14] about every component of the unknown density matrix ρ ,

$$\rho = \frac{1}{2} (\mathbb{1} + \langle \sigma_x \rangle \sigma_x + \langle \sigma_y \rangle \sigma_y + \langle \sigma_z \rangle \sigma_z). \quad (2)$$

The canonical example (which we consider hereafter) is to measure the three Pauli operators ($\sigma_x, \sigma_y, \sigma_z$). This minimizes the variance of the estimator $\hat{\rho}$ – but not the expected infidelity, for two reasons.

First, *the variance of the estimate $\hat{\rho}$ depends also on ρ itself*. Consider the linear inversion estimator $\hat{\rho}_{\text{lin}}$, defined by estimating $\langle \sigma_z \rangle = \frac{n_{\uparrow} - n_{\downarrow}}{n_{\uparrow} + n_{\downarrow}}$ (and similarly for $\langle \sigma_x \rangle$ and $\langle \sigma_y \rangle$), and substituting into Eq. 2. Each measurement behaves like $N/3$ flips of a coin with bias $p_k = \frac{1}{2}(1 + \langle \sigma_k \rangle)$, and yields

$$\hat{p}_k = p_k \pm \sqrt{\frac{3}{N}} \sqrt{p_k(1 - p_k)} \quad (3)$$

$$\Rightarrow \langle \sigma_k \rangle_{\text{estimated}} = \langle \sigma_k \rangle_{\text{true}} \pm \sqrt{\frac{3}{2N}} \sqrt{1 - \langle \sigma_k \rangle^2}. \quad (4)$$

When $\langle \sigma_k \rangle \approx 0$, its estimate has a large variance – but when $\langle \sigma_k \rangle \approx \pm 1$, the variance is very small. As a result, the variance of $\hat{\rho}$ around ρ is anisotropic and ρ -dependent (see Fig. 1a).

Second, *the dependence of infidelity on the error, $\Delta = \hat{\rho} - \rho$, also varies with ρ* . Infidelity is hypersensitive to

misestimation of small eigenvalues. A Taylor expansion of $1 - F(\rho, \hat{\rho})$ yields (in terms of ρ 's eigenbasis $\{|i\rangle\}$),

$$1 - F(\rho, \rho + \epsilon \Delta) = \frac{1}{4} \sum_{i,j} \frac{\langle i | \Delta | j \rangle^2}{\langle i | \rho | i \rangle + \langle j | \rho | j \rangle} + O(\Delta^3). \quad (5)$$

Infidelity is quadratic in Δ – except that as an eigenvalue $\langle i | \rho | i \rangle$ approaches 0, its sensitivity to variations in $\langle i | \Delta | i \rangle$ diverges, and $1 - F$ becomes *linear* [15] in Δ :

$$1 - F(\rho, \rho + \epsilon \Delta) = \epsilon \sum_{i: \langle i | \rho | i \rangle = 0} \langle i | \Delta | i \rangle + O(\Delta^2). \quad (6)$$

So to minimize infidelity, we must accurately estimate the small eigenvalues of ρ , particularly those that are (or appear to be) zero. For states deep within the Bloch sphere, static tomography achieves infidelity of $O(1/N)$ [9, 16]. Typical errors scale as $|\Delta| = O(1/\sqrt{N})$ (Eq. 4), and infidelity scales as $1 - F = O(|\Delta|^2)$. But for states with eigenvalues less than $O(1/\sqrt{N})$, infidelity scales as $O(1/\sqrt{N})$. Since quantum information processing relies on pure and nearly-pure states, it is this poor scaling (rather than the $O(1/N)$ scaling for highly mixed states) that is significant.

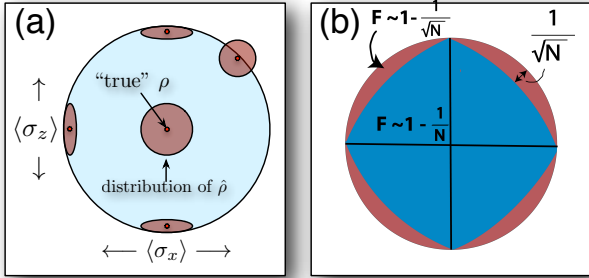


FIG. 1. Two features of qubit tomography with Pauli measurements (shown for an equatorial cross-section of the Bloch sphere): **(a)** The distribution or “scatter” of any unbiased estimator $\hat{\rho}$ (depicted by dull red ellipses) varies with the true state ρ (bright red dots at the center of ellipses); **(b)** The expected infidelity between $\hat{\rho}$ and ρ as a function of ρ . Within the Bloch sphere, the expected infidelity is $O(1/N)$. But in a thin shell of nearly-pure states (of thickness $O(1/\sqrt{N})$), it scales as $O(1/\sqrt{N})$ – *except* when ρ is aligned with a measurement axis (Pauli X, Y, or Z).

To achieve better performance, we observe that if ρ is diagonal in one of the measured bases (e.g., σ_z), then infidelity *always* scales as $O(1/N)$ – no matter what ρ 's eigenvalues are. The increased sensitivity of $1 - F$ to error in small eigenvalues (Eq. 5) is precisely canceled by the reduced inaccuracy that accompanies a highly biased measurement-outcome distribution (Eq. 4). This suggests an obvious (if naïve) solution: we should simply

choose one of our measurement bases to be the diagonal basis of ρ !

Of course, this is impossible – knowing ρ would render tomography pointless. But we *can* perform standard tomography on $N_0 < N$ samples, get a preliminary estimate $\hat{\rho}_0$, and measure the remaining $N - N_0$ samples in a frame where one basis diagonalizes $\hat{\rho}_0$. This measurement will not quite exactly diagonalize ρ , but if $N_0 \gg 1$ it will be fairly close. The angle θ between the eigenbases of ρ and $\hat{\rho}_0$ is $O(|\Delta|) = O(1/\sqrt{N_0})$. This implies that if ρ has an eigenvector $|\psi_k\rangle$ with eigenvalue $\lambda_k = 0$, then the probability of the corresponding measurement outcome $|\phi_k\rangle\langle\phi_k|$ will be at most $p_k = \sin^2 \theta \approx \theta^2 = O(1/N_0)$. Since we make this measurement on $O(N - N_0)$ copies [17], the final error in the estimated \hat{p}_k (and therefore in the eigenvalue λ_k) is $O(1/\sqrt{N_0(N - N_0)})$. So using a constant fraction $N_0 = \alpha N$ of the available samples for the preliminary estimation should yield $O(1/N)$ infidelity for *all* states.

A very similar protocol was suggested by Bagan *et al* in Ref. [11]. However, that analysis concluded that $N_0 \propto N^p$ for $p \geq \frac{2}{3}$ would be sufficient. This only applies to *average* infidelity when ρ is drawn from a particular ensemble. Our analysis shows that this choice yields $1 - F = O(N^{-5/6})$ for almost all nearly-pure states.

SIMULATION RESULTS

To confirm the theory, we did numerical simulations of single-qubit tomography using four different protocols: (1) standard fixed-measurement tomography; (2) adaptive tomography with $N_0 = N^{2/3}$, as proposed in [11]; (3) adaptive tomography with $N_0 = \alpha N$ (for a range of α); and (4) “known basis” tomography, wherein we cheat by adjusting our measurement frame for all N samples to align with ρ 's eigenbasis. We simulated many true states ρ , but present only the most interesting case, a pure state at 45 degrees to both the σ_x and σ_z axes. Our results are not particularly sensitive to the exact estimator used; we used maximum-likelihood estimation (MLE) with a quadratic approximation to the negative loglikelihood function:

$$l(\rho) = -\log \mathcal{L}(\rho) \approx \sum_{k=1}^3 \frac{N_k (\text{Tr}[\rho E_k] - f_k)^2}{f_k (1 - f_k)}, \quad (7)$$

where $f_k = n_k/N_k$ are the observed frequencies of the +1 eigenvectors of the three Pauli operators σ_k , E_k is the corresponding projector, and N_k is the number of samples on which σ_k was measured. Convex optimization (in MATLAB [18]) was used to find $\hat{\rho}_{\text{MLE}}$. Results were averaged over many (typically 150) randomly generated measurement records.

Figure 2 shows average infidelity versus N . We fit these simulated data to power laws of the form $1 - F = \beta N^p$,

and found $p = -0.513 \pm 0.006$ (for static tomography), $p = -0.868 \pm 0.008$ (for adaptive tomography with $N_0 = N^{2/3}$), $p = -0.980 \pm 0.006$ (for adaptive tomography with $N_0 = 0.5N$), and $p = -0.993 \pm 0.09$ (for known-basis tomography). These results are not significantly different [19] from predictions of the simple theory ($p = -\frac{1}{2}, -\frac{5}{6}, 1$, and 1 , respectively). The borderline-significant discrepancy is, we believe, due to boundary effects ($\hat{\rho}_{\text{MLE}}$ is constrained to be positive) that aren't modeled in the simple theory. We also varied $\alpha = N_0/N$ (Fig. 2, inset) and found that $\alpha = \frac{1}{2}$ optimizes the prefactor (β).

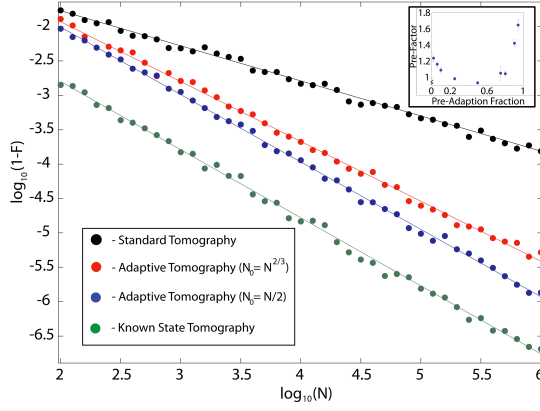


FIG. 2. Average infidelity $1 - F(\hat{\rho}, \rho)$ vs. sample size N for Monte Carlo simulations of four different tomographic protocols: standard tomography (black), the procedure proposed in [11] using $N_0 = N^{2/3}$ (red), our procedure using $N_0 = N/2$ (blue), and “known basis” tomography (green). Both adaptive procedures clearly outperform static tomography, but our procedure clearly outperforms the $N_0 = N^{2/3}$ approach, and matches the asymptotic scaling of known-basis tomography. The inset shows the dependence of the prefactor (β) on $\alpha = N_0/N$.

EXPERIMENTAL RESULTS

We implemented our simple adaptive protocol experimentally in linear optics. Using type-1 spontaneous parametric down conversion in a nonlinear crystal, photon pairs were created. One of these photons was sent immediately to a single photon counting module (SPCM) to act as a trigger. The second photon was sent through a Glan-Thomson polarizer to prepare it in a state of very pure linear polarization. Waveplates were first used to prepare the polarization state of the photon, and subsequently used in tandem with a polarization beamsplitter to project onto any state on the Bloch sphere. They are computer-controlled, enabling changes during the experiment.

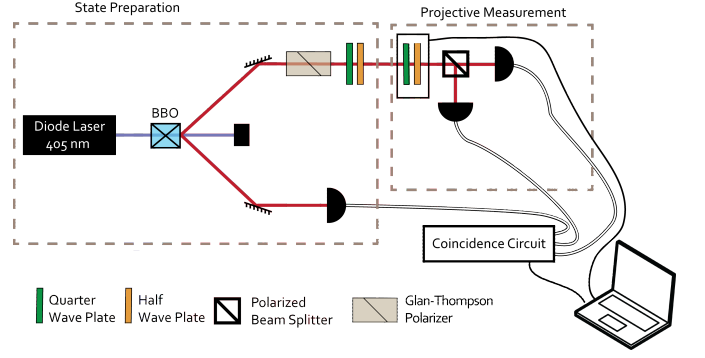


FIG. 3. Spontaneous parametric downconversion is performed by pumping a nonlinear BBO crystal with linearly polarized light. One photon is sent directly to a detector as a trigger. A rotation using a quarter-half waveplate combination prepares the other photon in any desired polarization state. Finally, a projective measurement onto any axis of the Bloch sphere is performed by a quarter-half waveplate combination followed by a polarizing beamsplitter. The measurement waveplates are connected to a computer to enable adaptation.

Our “standard tomography” protocol involved collecting $N/3$ photons at each of three measurement settings corresponding to σ_x , σ_y , and σ_z , and computing $\hat{\rho}_{\text{MLE}}$ as outlined in [20]. Each data point in figure 4 represents an average over many (~ 150) repetitions. The “true state” ρ was determined from one very long tomographic experiment in which $\tilde{N} = 10^7$ photons were collected. The overwhelming size of this dataset ensures accuracy sufficient to calibrate the other experiments, all of which involve $N \leq 3 \times 10^4$ photons.

To do adaptive tomography, we measure N_0 of the photons first and use the outcomes to generate an ML estimate $\hat{\rho}_0$. We then rotate the measurements so that one coincides with the eigenbasis of the preliminary estimate. Finally, we measure the remaining $N - N_0$ photons in this new set of bases and construct a final ML estimate of the state from all the data collected in both phases.

Figure 4 compares standard tomography to adaptive tomography with $N_0 = N^{2/3}$ and with $N_0 = N/2$. We fit a power law ($1 - F = \beta N^p$) to the average fidelity of each protocol, and found $p = -0.51 \pm 0.02$ for standard tomography, $p = -0.71 \pm 0.04$ for the procedure of Ref. [11], and $p = -0.90 \pm 0.04$ for our adaptive procedure.

Data confirm that adaptive tomography outperforms standard tomography by an order of magnitude even for modest ($\sim 10^4$) N . Our data generally agree with theory, but some experiments (specifically, those that achieve very low infidelities) show small but statistically significant discrepancies with theory. Infidelities as low as 10^{-4} are dominated by systematic errors not modeled in the theory. We believe the primary source of systematic errors is imperfections in the waveplates and their angles – in simulations, fluctuations on the order of 10^{-3} radians

are sufficient to reproduce the observed deviations.

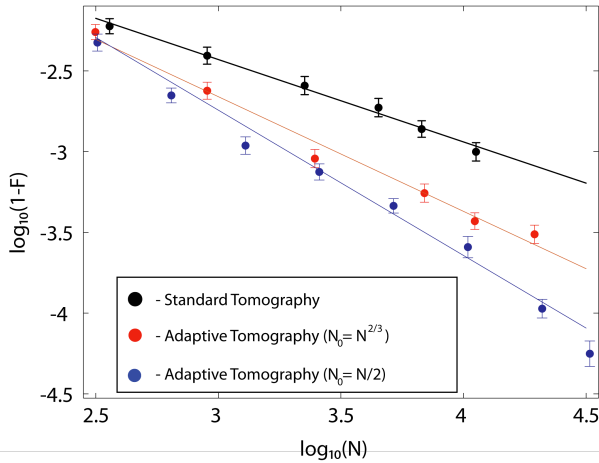


FIG. 4. Experimental data: The average infidelity $1 - F(\hat{\rho}, \rho)$ for the three tomographic protocols shown in Fig. 2 vs. the number of samples N . Each average is over 150 different realizations of the experiment; error bars are standard deviation of the mean of these samples.

Finally, we devised and performed an even simpler adaptive procedure. After using standard tomography on $N_0 = \frac{N}{2}$ samples, to get a preliminary estimate $\hat{\rho}_0$, we measured *all* of the remaining $N - N_0$ samples in the diagonal basis of $\hat{\rho}_0$. Fig. 5 shows the results; this *reduced adaptive tomography* procedure achieves the same $O(\frac{1}{N})$ infidelity. The best fits to the exponent p in $1 - F = \beta N^p$ are $p = -0.51 \pm 0.02$ for standard tomography and $p = -0.88 \pm 0.05$ for reduced adaptive tomography (not significantly different from the results shown in Fig. 4). Reduced adaptive tomography only requires one extra measurement setting (full adaptive tomography requires three). These procedures generalize to higher dimensional systems, where the advantage of reduced adaptive tomography over adaptive tomography is even greater.

DISCUSSION

We demonstrated two easily implemented adaptive tomography procedures that achieve $1 - F(\hat{\rho}, \rho) = O(1/N)$ for *every* qubit state. In contrast, any static tomography protocol will yield infidelity $O(1/\sqrt{N})$ for most nearly-pure states. These adaptive schemes require only marginally more resources than standard tomography (just one more measurement setting for the reduced scheme). We see almost no reason not to use reduced adaptive tomography in future experiments.

The $O(1/N)$ infidelity scaling achieved by our scheme is optimal, but the constant can surely be improved – i.e., if our scheme has asymptotic error α/N , a more sophis-

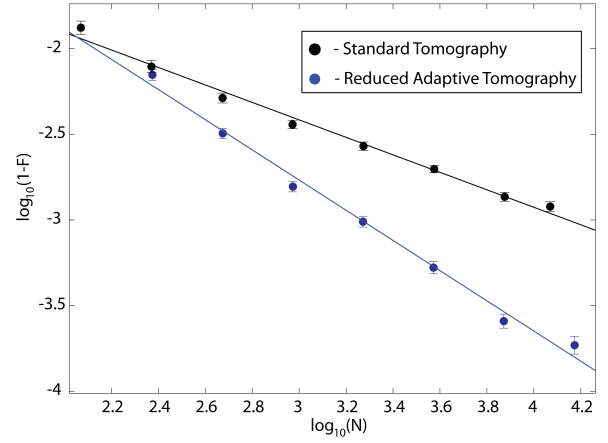


FIG. 5. Experimental data: standard tomography compared to *reduced* adaptive tomography. Average infidelity $1 - F(\hat{\rho}, \rho)$ for standard tomography (black) and reduced adaptive tomography (blue) is plotted versus N . Each average is over 200 different realizations of the experiment; error bars are standard deviation of the mean of these samples.

ticated scheme can achieve α'/N with $\alpha' < \alpha$. The absolutely optimal protocol requires joint measurements on all N samples [21], and while it still suffers error $O(1/N)$, this [unknown] protocol will outperform any local measurement. Even within local protocols, there is undoubtedly some marginal benefit to adapting more than once. What we have shown is that a single adaptation is sufficient for optimal worst-case scaling.

Closely related previous work [11] optimized average fidelity over a specific measure. They chose to average over Bures measure, a very respectable choice [22–24]. Unfortunately, the “hard-to-estimate” states lie in a thin shell at the surface of the Bloch sphere, and as $N \rightarrow \infty$, this shell’s Bures measure vanishes. So although the scheme with $N_0 \propto N^{2/3}$ proposed in [11] achieves Bures-average infidelity $O(1/N)$, it achieves only $O(1/N^{5/6})$ infidelity for nearly all of the (important) nearly-pure states.

Ironically, while the “hard-to-estimate” states are all nearly pure, restricting the problem to pure states (e.g., via a Bayesian prior supported *only* on pure states) falsely trivializes it – the average *and* worst-case risk drops to $O(1/N)$ even with static tomography! The critical difficulty is not in estimating which pure state we have (unitary errors don’t impact fidelity very much). It is in distinguishing between small eigenvalues – telling the difference between $\lambda = 0$ and $\lambda = 1/\sqrt{N}$.

We conclude with an observation that may surprise: adaptivity provides no advantage at all if inaccuracy is measured by trace-norm or 2-norm, which aren’t hypersensitive to small variations in small eigenvalues. This does *not* undermine our result – it has a simple explanation. Trace-norm ($\|\hat{\rho} - \rho\|_1$) quantifies *single-shot* distinguishability. When $N \gg 1$ samples are available, it be-

comes irrelevant. The relevant quantity is $|\hat{\rho}^{\otimes N} - \rho^{\otimes N}|_1$, whose behavior is defined by the Chernoff bound [25], which in turn is well approximated by infidelity. So infidelity is a measure of *many-copy* distinguishability. Since tomography is necessarily concerned with $N \gg 1$ copies, the advantages of adaptivity *are* real, and hold for all many-copy metrics (e.g., relative entropy, Chernoff bound, etc.)

DHM, LAR, AD, and AMS thank NSERC, CIFAR, FATR, CAF, and QuantumWorks for support. Additional thanks go to Alan Stummer for designing the coincidence circuit. RBK was supported by the LDRD program at Sandia National Laboratories, a multi-program laboratory operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

* dmahler@physics.utoronto.ca

- [1] Z. Hradil, Phys. Rev. A **55**, 1561 (1997).
- [2] M. Paris and J. Rehacek, *Quantum state estimation*, Vol. 649 (Springer, 2004).
- [3] R. Blume-Kohout, N. J. Phys. **12**, 043034 (2010).
- [4] R. Blume-Kohout, Phys. Rev. Lett. **105**, 200504 (2010).
- [5] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Phys. Rev. Lett. **105**, 150401 (2010).
- [6] M. Christandl and R. Renner, Phys. Rev. Lett. **109**, 120403 (2012).
- [7] R. Blume-Kohout, arXiv:1202.5270 (2012).
- [8] R. Okamoto, M. Iefuji, S. Oyama, K. Yamagata, H. Imai, A. Fujiwara, and S. Takeuchi, Phys. Rev. Lett. **109**, 130404 (2012).
- [9] R. D. Gill and S. Massar, Phys. Rev. A **61**, 042312 (2000).
- [10] E. Bagan, M. Ballester, R. D. Gill, A. Monras, and R. Muñoz-Tapia, Phys. Rev. A **73**, 032301 (2006).
- [11] E. Bagan, M. Ballester, R. Gill, R. Muñoz-Tapia, and O. Romero-Isart, Phys. Rev. Lett. **97**, 130501 (2006).
- [12] F. Huzár and N. M. T. Houlby, Phys. Rev. A **85**, 052120 (2012).
- [13] M. D. de Burgh, N. K. Langford, A. C. Doherty, and A. Gilchrist, Phys. Rev. A **78**, 052122 (2008).
- [14] A. J. Scott, J. Phys. A **39**, 13507 (2006).
- [15] Because ρ lies on the state-set's boundary, the gradient of F need not vanish in order for $\hat{\rho} = \rho$ to be a local maximum.
- [16] T. Sugiyama, P. S. Turner, and M. Murao, N. J. Phys. **14**, 085005 (2012).
- [17] The “ O ” notation is necessary here because some of the remaining $N - N_0$ copies may be measured in other bases that make up a complete measurement frame.
- [18] J. Löfberg, Proc. CACSD (Taipei) (2004).
- [19] All quoted uncertainties herein are 1σ , or 68% confidence intervals. Therefore, we don't expect the “true” value to lie within the error bars more than 68% of the time. Most of the results given here agree with theoretical predictions to within 2σ (95% confidence intervals), a common criterion for consistency between data and theory.
- [20] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Phys. Rev. A **64**, 052312 (2001).
- [21] S. Massar and S. Popescu, Phys. Rev. Lett. **74**, 1259 (1995).
- [22] M. Hübner, Phys. Lett. A **163**, 239 (1992).
- [23] D. Petz and C. Sudar, J. Math. Phys. **37**, 2662 (1996).
- [24] K. Życzkowski and H.-J. Sommers, Phys. Rev. A **71**, 032313 (2005).
- [25] K. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, L. Masanes, A. Acín, and F. Verstraete, Phys. Rev. Lett. **98**, 160501 (2007).